DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A BAYESIAN FACTOR ANALYSIS MODEL WITH GENERALIZED PRIOR INFORMATION

Daniel B. Rowe

SOCIAL SCIENCE WORKING PAPER 1099

August 2000

A Bayesian Factor Analysis Model With Generalized Prior Information

Daniel B. Rowe
Division of Humanities and Social Sciences,
California Institute of Technology, USA

Abstract

In the Bayesian approach to factor analysis, available prior knowledge regarding the model parameters is quantified in the form of prior distributions and incorporated into the inferences. The incorporation of prior knowledge has the added consequence of eliminating the ambiguity of rotation found in the traditional factor analysis model. Previous Bayesian factor analysis work (Press & Shigemasu 1989, & Press 1998, Rowe 2000a, and Rowe 2000b), has considered mainly natural conjugate prior distributions for the model parameters. As is mentioned in Press (1982), Rothenburg (1963) pointed out that the with a natural conjugate prior distribution, the elements in the covariance matrices are constrained and thus may not be rich enough to permit freedom of assessment. In this paper, generlized natural conjugate distributions are used to quantify and incorporate available prior information which permit complete freedom of assessment.

1 Introduction

A factor analysis is performed to explain the relationship among a set of observed variables in terms of a smaller number of unobserved variables or latent factors which underlie the observations. This smaller number of variables can be used to find a meaningful structure in the observed variables. This structure will aid in the interpretation and explanation of the process that has generated the observations.

In the Bayesian approach to factor anlysis (Press & Shigemasu 1989, Rowe & Press 1998, Rowe 2000a, and Rowe 2000b henceforth PS89, RP98, R00a, and R00b) the classical normal sampling model is assumed, but the disturbance covariance matrix is specified to be a full positive definite matrix. One of the prior assumptions (PS89) is that the prior expected value of the disturbance covariance matrix is diagonal in order to represent traditional views of the factor model containing "common" and "specific" factors. As noted in Press (1982), Rothenburg (1963) pointed out that the natural conjugate prior distribution has covariance matrix elements that are constrained and thus may not be rich enough to assess the prior parameters. In this paper, generalized natural conjugate prior distributions are specified for the unknown matrices which permit complete freedom of assessment.

Bayesian statistical methods not only incorporate available prior information either from substantive experts or previous data, but allow the knowledge regarding the parameter values to accumulate as subsequent data is acquired. In the non-Bayesian factor analysis model, the factor loading matrix is determinate up to an orthogonal rotation. Typically after a non-Bayesian factor analysis, an orthogonal rotation is performed on the factor loading matrix according to one of many subjective criteria. This is not the case in Bayesian factor analysis. The rotation is automatically found.

The model parameters are estimated by both Gibbs sampling (Geman & Geman, 1984 and Gelfand & Smith, 1990) and iterated conditional modes (Lindley & Smith, 1972 and O'Hagen, 1994) which find posterior marginal mean and posterior joint modal (maximum a posteriori) estimates respectively.

The plan of the paper is to review the model and to adopt prior distributions in Section 2. Present the conditional posterior distributions along with the Gibbs sampling and ICM algorithms in Section 3. In Section 4 an example is detailed, and estimates from both the Gibbs sampling and the ICM estimation methods are presented.

2 Model

2.1 Likelihood Function

The Bayesian factor analysis model is:

$$\underset{(p\times 1)}{(x_j|\mu,\Lambda,f_j)} = \underset{(p\times 1)}{\mu} + \underset{(p\times m)}{\Lambda}\ \underset{(m\times 1)}{f_j} + \underset{(p\times 1)}{\epsilon_j}, \quad m<p, \tag{2.1}$$

for $j=1,\ldots,n$, where x_j is the j^{th} observation, μ is the overall population mean, Λ is a matrix of constants called the factor loading matrix; f_j is the factor score vector for subject j; and the ϵ_j's are assumed to be mutually uncorrelated and normally distributed $N(0, \Psi)$ variables.

In the traditional model, Ψ is taken to be a diagonal matrix so that common and specific factors can readily distinguished. The Bayesian models, take Ψ to be a general symmetric, positive definite covariance matrix with the property of being a priori diagonal on the average, i.e., $E(\Psi)$ = a diagonal matrix.

It is assumed that μ, Λ, the f_i's, and Ψ are unobservable and that the distribution of each x_j can be written as

$$p(x_j|\mu, \Lambda, f_j, \Psi) = (2\pi)^{-\frac{p}{2}}|\Psi|^{-\frac{1}{2}}e^{-\frac{1}{2}(x_j-\mu-\Lambda f_j)'\Psi^{-1}(x_j-\mu-\Lambda f_j)}. \tag{2.2}$$

If proportionality is denoted by "$\propto$" and the Kroneker product by $\otimes$ then, the likelihood for (μ, Λ, F, Ψ) is

$$p(X|\mu, \Lambda, F, \Psi) \propto |\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')} \tag{2.3}$$

where the p-variate observation vectors on n subjects are, $X' = (x_1, \ldots, x_n)$, the factor scores are $F' = (f_1, \ldots, f_n)$, and the errors of observation are $E' = (\epsilon_1, \ldots, \epsilon_n)$. The notation $p(\cdot)$ will generically denote a distribution which is distinguished by its argument. The proportionality constant in (2.3) depends only on (p, n) and not on (μ, Λ, F, Ψ).

2.2 Priors

Generalized natural conjugate prior distributions are specified for the unknown parameters which permit complete freedom of assessment. The joint prior distribution is:

$$p(\mu,\Lambda,F,\Psi) \propto p(\mu)p(\Lambda)p(\Psi)p(F), \tag{2.4}$$

where

$$\begin{aligned} p(\mu) &\propto |\Gamma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_0)'\Gamma^{-1}(\mu-\mu_0)} &(2.5)\\ p(\lambda) &\propto |\Delta|^{-\frac{1}{2}} e^{-\frac{1}{2}(\lambda-\lambda_0)'\Delta^{-1}(\lambda-\lambda_0)}, &(2.6)\\ p(\Psi) &\propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2}tr\Psi^{-1}B}, \quad \nu > 2p, &(2.7)\\ p(F) &\propto e^{-\frac{1}{2}trF'F} &(2.8) \end{aligned}$$

with $\Gamma, \Delta, B, \Psi > 0$ and B a diagonal matrix. A generalized natural conjugate normal distribution is specified for the population mean where μ_0 and Γ are mean and covariance hyperparameters to be assessed. The vector $\lambda = vec(\Lambda')$ is specified to have the generalized natural conjugate normal distribution with mean and covariance hyperparameters $\lambda_0 = vec(\Lambda_0')$ and Δ. The matrix Ψ follows an Inverted Wishart distribution, with hyperparameters (ν, B) which are to be assessed. It is assumed that $E(\Psi)$ is a priori diagonal, in order to represent traditional psychometric views of the factor model containing "common" and "specific" factors.

2.3 Joint Posterior

Using Bayes rule, combine (2.3)–(2.8), to get the joint posterior density of the parameters

$$\begin{aligned} p(\mu,F,\Lambda,\Psi|X) \propto\ & e^{-\frac{1}{2}trF'F}|\Gamma|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mu-\mu_0)'\Gamma^{-1}(\mu-\mu_0)}|\Delta|^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda-\lambda_0)'\Delta^{-1}(\lambda-\lambda_0)}\\ & \times|\Psi|^{-\frac{(n+\nu)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}[(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')+B]} \quad (2.9) \end{aligned}$$

3 Estimation

3.1 Conditional Posterior Densities

Both the Gibbs sampling and ICM estimation procedures require the posterior conditional distributions. Gibbs sampling requires the conditionals for the generation of random samples while ICM requires them for maximization by cycling through their modes.

The conditional posterior distribution of the overall mean is

$$\begin{aligned} p(\mu|F,\Lambda,\Psi,X) &\propto p(\mu)p(X|\mu,F,\Lambda,\Psi) \\ &\propto e^{-\frac{1}{2}(\mu-\tilde{\mu})'\left[(n\Gamma)^{-1}+\Psi^{-1}\right](\mu-\tilde{\mu})} \end{aligned} \tag{3.1}$$

where

$$\tilde{\mu} = \left[(n\Gamma)^{-1} + \Psi^{-1}\right]^{-1}\left[(n\Gamma)^{-1}\mu_0 + \Psi^{-1}(\bar{x} - \Lambda\bar{f})\right] \tag{3.2}$$

The mean of the factor scores has been denoted by $\bar{f}$. The overall mean given the factor scores, the factor loadings, the disturbance covariance matrix, and the data is normally distributed.

The conditional posterior distribution of the factor scores is

$$\begin{aligned} p(F|\mu,\Lambda,\Psi,X) &\propto p(F)p(X|\mu,F,\Lambda,\Psi) \\ &\propto e^{-\frac{1}{2}tr(F-\tilde{F})(I_m+\Lambda'\Psi^{-1}\Lambda)(F-\tilde{F})'} \end{aligned} \tag{3.3}$$

where $\tilde{F} \equiv (X - e_n \otimes \mu')\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$. The factor scores given the overall mean, the factor loadings, the disturbance covariance matrix, and the data is normally distributed.

The conditional posterior distribution of the factor loadings is

$$p(\Lambda|\mu,F,\Psi,X) \propto p(\lambda)p(X|\mu,F,\Lambda,\Psi)$$

$$\propto \quad e^{-\frac{1}{2}(\lambda-\tilde{\lambda})[\Delta^{-1}+\Psi^{-1}\otimes F'F](\lambda-\tilde{\lambda})'} \tag{3.4}$$

where

$$\tilde{\lambda} = [\Delta^{-1} + \Psi^{-1} \otimes F'F]^{-1}[\Delta^{-1}\lambda_0 + (\Psi^{-1} \otimes F'F)\lambda_\star] \tag{3.5}$$

and

$$\lambda_\star = vec[(F'F)^{-1}F'(X - e_n \otimes \tilde{\mu}')] \tag{3.6}$$

The conditional posterior density of the factor loadings given the overall mean, the factor scores, the disturbance covariance matrix, and the data is normally distributed.

The conditional posterior distribution of the disturbance covariance matrix is

$$\begin{aligned} p(\Psi|\mu, F, \Lambda, X) &\propto p(\Psi)p(X|\mu, F, \Lambda, \Psi) \\ &\propto |\Psi|^{-\frac{(n+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}[(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')+B]}. \end{aligned} \tag{3.7}$$

The conditional density of the disturbance covariance matrix given the overall mean, the factor scores, the factor loadings, and the data has an inverted Wishart density.

The modes of these conditional distributions are $\tilde{\mu}$, $\tilde{F}$, $\tilde{\lambda}$ (as defined above), and

$$\tilde{\Psi} = \frac{(X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda') + B}{n + \nu}, \tag{3.8}$$

respectively.

3.2 The Gibbs Sampling Algorithm

For Gibbs sampling estimation of the posterior, start with $\bar{\mu}_{(0)}$, $\bar{F}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned}
\bar{\Lambda}_{(l+1)} &= \text{a random sample from } p(\Lambda|\bar{\mu}_{(l)}, \bar{F}_{(l)}, \bar{\Psi}_{(l)}, X) \\
\bar{\Psi}_{(l+1)} &= \text{a random sample from } p(\Psi|\bar{\mu}_{(l)}, \bar{F}_{(l)}, \bar{\Lambda}_{(l+1)}, X) \\
\bar{F}_{(l+1)} &= \text{a random sample from } p(F|\bar{\mu}_{(l)}, \bar{\Lambda}_{(l+1)}, \bar{\Psi}_{(l+1)}, X) \\
\bar{\mu}_{(l+1)} &= \text{a random sample from } p(\mu|\bar{F}_{(l+1)}, \bar{\Lambda}_{(l+1)}, \bar{\Psi}_{(l+1)}, X)
\end{aligned}$$

and after the first random samples called the "burn in" are discarded compute from the next L samples

$$\bar{F} = \frac{1}{L}\sum_{l=1}^{L}\bar{F}_{(l)} \qquad \bar{\Lambda} = \frac{1}{L}\sum_{l=1}^{L}\bar{\Lambda}_{(l)} \qquad \bar{\Psi} = \frac{1}{L}\sum_{l=1}^{L}\bar{\Psi}_{(l)} \qquad \bar{\mu} = \frac{1}{L}\sum_{l=1}^{L}\bar{\mu}_{(l)}$$

which are the sampling based marginal posterior mean and modal estimates of the parameters.

3.3 The ICM Algorithm

For the ICM estimation of the parameters start with initial values for $\tilde{\mu}$ $\tilde{F}$, and $\tilde{\Psi}$ say $\tilde{\mu}_{(0)}$, $\tilde{F}_{(0)}$, and $\tilde{\Psi}_{(0)}$ then cycle through

$$\begin{aligned}
\tilde{\lambda}_{(l+1)} &= [\Delta^{-1} + \Psi_{(l)}^{-1} \otimes F'_{(l)}F_{(l)}]^{-1} \\
&\quad \cdot \left\{\Delta^{-1}\lambda_0 + (\Psi_{(l)}^{-1} \otimes F'_{(l)}F_{(l)})vec[(F'_{(l)}F_{(l)})^{-1}F'_{(l)}(X - e_n \otimes \tilde{\mu}'_{(l)})]\right\} \\
\tilde{\Psi}_{(l+1)} &= \frac{(X - e_n \otimes \tilde{\mu}'_{(l)} - \tilde{F}_{(l)}\tilde{\Lambda}'_{(l+1)})'(X - e_n \otimes \tilde{\mu}'_{(l)} - \tilde{F}_{(l)}\tilde{\Lambda}'_{(l+1)}) + B}{n+\nu} \\
\tilde{F}_{(l+1)} &= (X - e_n \otimes \tilde{\mu}'_{(l)})\tilde{\Psi}_{(l+1)}^{-1}\tilde{\Lambda}_{(l+1)}(I_m + \tilde{\Lambda}'_{(l+1)}\tilde{\Psi}_{(l+1)}^{-1}\tilde{\Lambda}_{(l+1)})^{-1}. \\
\tilde{\mu}_{(l+1)} &= \left[(n\Gamma)^{-1} + \Psi_{(l+1)}^{-1}\right]^{-1}\left[(n\Gamma)^{-1}\mu_0 + \Psi_{(l+1)}^{-1}(\bar{x} - \tilde{\Lambda}_{(l+1)}\bar{\tilde{f}}_{(l+1)})\right]
\end{aligned}$$

until convergence is reached with the joint posterior modal estimators also known as the maximum a posteriori estimators for the unknown parameters $(\tilde{\mu}, \tilde{F}, \tilde{\Lambda}, \tilde{\Psi})$.

4 Example

In this section the Gibbs sampling and the ICM procedures for estimating the parameters of the Bayesian factor analysis model are implemented and the resulting estimators are presented. The data is extracted from an example in Kendall 1980, p.53. The problem as originally stated in PS89 and again in the subsequent Bayesian factor analysis papers is the following.

> There are 48 applicants for a certain job, and they have been scored on 15 variables regarding their acceptability. They are:

(1)	Form of letter application	(9)	Experience
(2)	Appearance	(10)	Drive
(3)	Academic ability	(11)	Ambition
(4)	Likeabiliy	(12)	Grasp
(5)	Self-confidence	(13)	Potential
(6)	Lucidity	(14)	Keenness to join
(7)	Honesty	(15)	Suitability
(8)	Salesmanship		

The raw scores of the applicants on these 15 variables, measured on the same scale, are presented in Table 1. The question is, Is there an underlying subset of factors that explain the variation observed in the scores? If so, then the applicants could be compared more easily.

The same underlying structure is postulated as in as PS89, a model with 4 factors. This choice is based upon PS89 having carried out a principal

Table 1: Raw scores of 48 applicants scaled on 15 variables.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	6	7	2	5	8	7	8	8	3	8	9	7	5	7	10
2	9	10	5	8	10	9	9	10	5	9	9	8	8	8	10
3	7	8	3	6	9	8	9	7	4	9	9	8	6	8	10
4	5	6	8	5	6	5	9	2	8	4	5	8	7	6	5
5	6	8	8	8	4	5	9	2	8	5	5	8	8	7	7
6	7	7	7	6	8	7	10	5	9	6	5	8	6	6	6
7	9	9	8	8	8	8	8	8	10	8	10	8	9	8	10
8	9	9	9	8	9	9	8	8	10	9	10	9	9	9	10
9	9	9	7	8	8	8	8	5	9	8	9	8	8	8	10
10	4	7	10	2	10	10	7	10	3	10	10	10	9	3	10
11	4	7	10	0	10	8	3	9	5	9	10	8	10	2	5
12	4	7	10	4	10	10	7	8	2	8	8	10	10	3	7
13	6	9	8	10	5	4	9	4	4	4	5	4	7	6	8
14	8	9	8	9	6	3	8	2	5	2	6	6	7	5	6
15	4	8	8	7	5	4	10	2	7	5	3	6	6	4	6
16	6	9	6	7	8	9	8	9	8	8	7	6	8	6	10
17	8	7	7	7	9	5	8	6	6	7	8	6	6	7	8
18	6	8	8	4	8	8	6	4	3	3	6	7	2	6	4
19	6	7	8	4	7	8	5	4	4	2	6	8	3	5	4
20	4	8	7	8	8	9	10	5	2	6	7	9	8	8	9
21	3	8	6	8	8	8	10	5	3	6	7	8	8	5	8
22	9	8	7	8	9	10	10	10	3	10	8	10	8	10	8
23	7	10	7	9	9	9	10	10	3	9	9	10	9	10	8
24	9	8	7	10	8	10	10	10	2	9	7	9	9	10	8
25	6	9	7	7	4	5	9	3	2	4	4	4	4	5	4
26	7	8	7	8	5	4	8	2	3	4	5	6	5	5	6
27	2	10	7	9	8	9	10	5	3	5	6	7	6	4	5
28	6	3	5	3	5	3	5	0	0	3	3	0	0	5	0
29	4	3	4	3	3	0	0	0	0	4	4	0	0	5	0
30	4	6	5	6	9	4	10	3	1	3	3	2	2	7	3
31	5	5	4	7	8	4	10	3	2	5	5	3	4	8	3
32	3	3	5	7	7	9	10	3	2	5	3	7	5	5	2
33	2	3	5	7	7	9	10	3	2	2	3	6	4	5	2
34	3	4	6	4	3	3	8	1	1	3	3	3	2	5	2
35	6	7	4	3	3	0	9	0	1	0	2	3	1	5	3
36	9	8	5	5	6	6	8	2	2	2	4	5	6	6	3
37	4	9	6	4	10	8	8	9	1	3	9	7	5	3	2
38	4	9	6	6	9	9	7	9	1	2	10	8	5	5	2
39	10	6	9	10	9	10	10	10	10	10	8	10	10	10	10
40	10	6	9	10	9	10	10	10	10	10	10	10	10	10	10
41	10	7	8	0	2	1	2	0	10	2	0	3	0	0	10
42	10	3	8	0	1	1	0	0	10	0	0	0	0	0	10
43	3	4	9	8	2	4	5	3	6	2	1	3	3	3	8
44	7	7	7	6	9	8	8	6	8	8	10	8	8	6	5
45	9	6	10	9	7	7	10	2	1	5	5	7	8	4	5
46	9	8	10	10	7	9	10	3	1	5	7	9	9	4	4
47	0	7	10	3	5	0	10	0	0	2	2	0	0	0	0
48	0	6	10	1	5	0	10	0	0	2	2	0	0	0	0

components analysis and having found that 4 factors accounted for 81.5% of the variance. Based upon underlying theory they constructed the prior factor loading matrix

$$\Lambda_0' = \begin{pmatrix} 0 & 0 & 0 & 0 & .7 & .7 & 0 & .7 & 0 & .7 & .7 & .7 & .7 & 0 & 0 \\ 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & .7 \\ 0 & 0 & 0 & .7 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The hyperparameters were assessed as $\mu_0 = 7.5e_{15}$, $\Gamma = \gamma_0 I_{15}$ where $\gamma_0 = 1/25$, $\Delta = \delta_0 I_{60}$ where $\delta_0 = 1/100$, $B = 0.2I_{15}$, and $\nu = 33$. The 15 dimensional unit vector has been denoted by e_{15}. The population mean, factor scores, factor loadings, and disturbance covariance matrix may now be estimated. It was found that a burn in period of $5,000$ samples worked well, so then the next $25,000$ samples were taken for the Gibbs estimates.

Table 2 displays the Gibbs sampling and ICM estimates of the population mean along with the prior and sample means.

Table 2: Gibbs Sampling and ICM estimates of the mean.

p	Gibbs Mean	ICM Mean	Sample Mean	Prior Mean
1	6.1739	5.9391	6.0000	7.5000
2	7.1735	7.0027	7.0833	7.5000
3	6.9847	7.1059	7.0833	7.5000
4	6.2971	6.1951	6.1458	7.5000
5	6.9259	6.9077	6.9375	7.5000
6	6.2857	6.3092	6.3333	7.5000
7	7.9717	7.9915	8.0417	7.5000
8	4.8728	4.8341	4.7917	7.5000
9	4.2777	4.3010	4.2292	7.5000
10	5.4062	5.3594	5.3125	7.5000
11	6.1025	5.9994	5.9792	7.5000
12	6.2211	6.2058	6.2500	7.5000
13	5.7116	5.6898	5.6875	7.5000
14	5.8657	5.6394	5.5625	7.5000
15	5.9462	5.9141	5.9583	7.5000

Table 3 displays the Gibbs sampling and ICM estimates of the factor loadings. For enhanced interpretability, the rows of the factor loading matrices have been rearranged. It is seen that factor 1 loads heavilly for variables 5, 6, 8, 10, 11, 12, and 13; factor 2

Table 3: Gibbs (left) and ICM (right) Estimates of Factor Loadings.

p	1	2	3	4	1	2	3	4
5	0.7566	-0.0066	-0.1438	-0.0084	0.8320	-0.0655	-0.1939	-0.0136
6	0.7699	-0.0604	-0.0256	0.0919	0.8134	-0.0013	-0.0509	0.0601
8	0.7481	-0.0640	0.0597	-0.0822	0.8057	-0.0637	0.0520	-0.0972
10	0.6818	0.0307	0.1528	-0.0101	0.7263	-0.0081	0.1582	-0.0053
11	0.7582	0.0253	-0.0245	-0.1161	0.8333	-0.0583	-0.0212	-0.1260
12	0.7074	0.0271	0.0915	0.0992	0.7454	0.0959	0.0896	0.0913
13	0.6600	0.1183	0.1235	0.1954	0.6858	0.1983	0.1273	0.2024
3	0.1012	0.6236	0.0667	0.0222	0.0972	0.9825	0.0474	0.0201
1	0.0118	0.0849	0.6633	-0.0230	-0.0033	-0.0666	0.8054	-0.0056
9	0.0018	0.0714	0.8134	-0.0363	-0.0240	0.1304	0.8670	-0.0606
15	0.2054	-0.1214	0.7165	0.0084	0.1885	-0.0164	0.7807	0.0007
4	0.0940	0.0086	0.0980	0.7307	0.0575	0.0034	0.1298	0.8262
7	0.0771	-0.0046	-0.1385	0.7399	0.0541	0.0166	-0.1953	0.8689
2	0.1819	0.0037	0.0971	0.1287	0.2071	0.0401	0.1362	0.2127
14	0.2658	-0.1010	0.1585	0.2967	0.2839	-0.3710	0.2221	0.3844

heavily on variable 3; factor 3 heavily on variables 1, 9, and 15; while factor 4 loads heavily on variables 4 and 7. These factors in terms of the original variables are factor 1: Self-confidence, Lucidity, Salesmanship, Drive, Ambition, Grasp, Potential; factor 2: Academic ability; factor 3: Form of letter application, Experience, Suitability; and factor 4: Likeabiliy, Honesty. These factors may be loosely interpreted as factor 1 being personality, factor 2 being academic ability, factor 3 being position match, and factor 4 being charisma.

Table 4: Gibbs (left) and ICM (right) Estimates of the Factor Scores.

Person	1	2	3	4	1	2	3	4
1	0.7140	-1.9888	0.2867	-0.6218	0.6951	-2.5776	0.2752	-0.4160
2	1.3259	-0.7555	0.8793	0.2929	1.3230	-1.2788	0.9822	0.5433
3	0.9309	-1.4508	0.5356	-0.1136	0.9411	-2.1687	0.5482	0.1216
4	-0.3405	0.6089	0.4310	0.2253	-0.3081	0.6412	0.2213	0.1491
5	-0.3696	0.4571	0.8708	0.9623	-0.3562	0.6182	0.8069	0.8835
6	0.1306	0.1868	0.9106	0.4239	0.1676	0.0581	0.6542	0.3979
7	1.0077	0.4652	1.7056	0.2322	1.0227	0.2124	1.6539	0.3739
8	1.2818	0.7987	1.6796	0.2598	1.2988	0.5295	1.6487	0.4438
9	0.7080	0.1962	1.5633	0.3329	0.7517	-0.2185	1.5417	0.4506
10	1.8737	0.0391	-0.0453	-1.4510	1.7052	1.3982	-0.2289	-1.3233
11	1.5327	1.0274	-0.3984	-2.7489	1.3976	1.5924	-0.4641	-2.5343
12	1.5441	0.3806	-0.5856	-0.7435	1.3780	1.5545	-0.6357	-0.7939
13	-0.5857	0.0981	0.3231	1.2089	-0.5408	0.4644	0.3879	1.1273
14	-0.6899	0.7296	0.3509	0.7390	-0.6002	0.6032	0.4900	0.6784
15	-0.7381	0.1722	0.4033	0.8408	-0.7024	0.8409	0.1375	0.7050
16	0.8283	-0.8845	1.1461	0.0746	0.7863	-0.4897	0.9355	0.1083
17	0.2957	0.3286	0.7086	-0.0582	0.3747	-0.2965	0.7020	0.0716
18	-0.1037	0.2229	-0.5648	-0.9438	-0.0650	0.1384	-0.4250	-0.7719
19	-0.1101	0.1776	-0.3913	-1.0669	-0.1268	0.3389	-0.2940	-1.0470
20	0.7411	-0.6651	-0.2223	1.0285	0.6988	-0.1998	-0.1926	0.9932
21	0.5740	-0.9453	-0.2811	0.9114	0.5225	-0.2798	-0.4012	0.7514
22	1.4964	0.0304	0.3143	0.6745	1.4311	-0.5712	0.5679	0.9006
23	1.4942	-0.0733	0.0916	0.9081	1.4554	-0.4807	0.3525	1.1775
24	1.3049	-0.0803	0.1912	1.2605	1.2111	-0.5076	0.5209	1.3426
25	-0.8499	0.0252	-0.5236	0.4718	-0.8237	0.0383	-0.3453	0.5248
26	-0.6843	0.1216	-0.0256	0.4909	-0.6491	0.0735	0.1263	0.4427
27	0.3346	-0.6543	-0.7885	1.1147	0.3022	0.2339	-0.8666	0.9531
28	-1.6430	0.1659	-1.4207	-1.3390	-1.5624	-1.1285	-1.1765	-1.3219
29	-1.7803	-0.0465	-1.5738	-2.4565	-1.7351	-1.5345	-1.1819	-2.3758
30	-0.9408	-0.4389	-1.1083	0.3998	-0.7865	-1.2336	-1.1128	0.4581
31	-0.6097	-0.3236	-0.8414	0.5941	-0.4939	-1.6605	-0.7658	0.6162
32	-0.0884	-0.7433	-1.1692	0.9123	-0.1854	-0.7180	-1.2705	0.4763
33	-0.3806	-1.0130	-1.3048	0.9618	-0.4488	-0.7622	-1.4646	0.4756
34	-1.4106	-0.1984	-1.2269	-0.3084	-1.3970	-0.4254	-1.1904	-0.3817
35	-2.0606	-0.6004	-0.7431	-0.3883	-1.9216	-1.3276	-0.6307	-0.1982
36	-0.6739	-0.0169	-0.4273	-0.0641	-0.6279	-0.9088	-0.1325	0.0285
37	0.6914	-0.4223	-1.4695	-0.8699	0.6650	-0.4001	-1.3759	-0.7181
38	0.7829	-0.4573	-1.4929	-0.5310	0.7348	-0.5774	-1.2481	-0.4344
39	1.5016	0.7616	1.8488	1.2583	1.4444	0.4937	1.7576	1.1872
40	1.6267	0.8545	1.7932	1.1504	1.5774	0.4579	1.7306	1.1107
41	-2.1533	-0.0585	2.1478	-2.5808	-2.0795	0.8354	1.8188	-2.5170
42	-2.4959	-0.0918	2.1325	-2.9032	-2.4608	0.7578	1.7721	-3.0978
43	-1.4252	-0.4100	0.4625	0.1542	-1.4961	1.1645	0.1847	-0.3606
44	0.9199	0.6363	0.4559	-0.2566	0.9209	-0.0114	0.4074	-0.1977
45	-0.0568	1.2599	-0.3682	1.2512	-0.0976	1.5153	-0.1588	0.9616
46	0.4545	1.2788	-0.5702	1.4274	0.3728	1.5767	-0.2367	1.1710
47	-2.0000	0.8790	-1.9891	-0.3394	-1.8717	1.7408	-2.1807	-0.3298
48	-1.9861	0.8993	-1.9951	-0.7853	-1.8642	1.7181	-2.2376	-0.7262

In Table 4, the Gibbs sampling and ICM estimates of the factor scores are presented. Note the similarity of most of the values but there are some minor differences.

Table 5: Gibbs (top) and ICM (bottom) Estimates of the Disturbance Covariance Matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.2831	.0309	-.1069	.0537	.0098	.0066	-.0316	-.0005	-.0712	-.0122	.0341	.0018	-.0102	.1307	-.0316
2		.5393	.0281	.0181	.0222	-.0601	.0583	.0303	-.0293	-.0638	.0912	.0192	.0122	-.0369	.0429
3			.5121	-.0362	-.0460	.0078	-.0119	-.0164	.0350	-.0263	-.0550	.0364	.0657	-.2456	.0226
4				.1098	-.0235	-.0002	-.0592	.0105	-.0153	-.0031	.0295	-.0178	.0044	.0643	.0029
5					.1251	-.0064	.0583	.0087	.0112	-.0043	.0202	-.0293	-.0321	.0140	-.0169
6						.0898	-.0273	-.0078	.0068	-.0417	-.0420	.0275	-.0172	-.0264	-.0120
7							.1624	.0097	.0051	.0243	-.0022	-.0031	-.0160	-.0119	.0042
8								.0812	-.0074	-.0001	.0054	-.0320	-.0277	.0249	.0091
9									.0980	-.0117	.0016	.0068	.0043	-.0256	-.0382
10										.1380	-.0018	-.0446	.0037	.0526	.0130
11											.1032	-.0140	-.0007	.0520	-.0095
12												.0929	.0163	-.0268	-.0041
13													.0901	-.0510	.0044
14														.3246	-.0327
15															.0966

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.1747	.0018	.0215	.0074	.0187	.0100	-.0168	-.0187	-.0879	-.0286	.0149	.0070	.0001	.0245	-.0783
2		.4012	-.0121	-.0353	.0140	-.0570	.0065	.0215	-.0410	-.0705	.0682	-.0020	-.0186	-.0665	.0246
3			.0220	.0031	.0054	-.0085	-.0085	.0091	-.0161	.0085	.0058	-.0147	-.0178	.0368	-.0110
4				.0806	-.0226	.0205	-.0711	.0050	-.0180	-.0122	.0119	-.0093	.0094	.0074	-.0098
5					.0816	-.0137	.0338	-.0052	.0246	-.0148	.0027	-.0272	-.0255	-.0102	-.0118
6						.0965	-.0173	-.0074	.0049	-.0452	-.0460	.0379	-.0096	-.0156	-.0035
7							.0867	.0056	.0316	.0172	-.0116	-.0045	-.0178	-.0324	.0199
8								.0721	-.0033	-.0024	-.0097	-.0316	-.0254	-.0028	.0214
9									.1262	-.0048	-.0017	.0005	-.0078	-.0056	-.0245
10										.1278	-.0134	-.0478	.0051	.0224	.0202
11											.0755	-.0170	.0004	.0080	-.0230
12												.0876	.0096	-.0084	-.0088
13													.0668	-.0106	-.0050
14														.1079	-.0471
15															.1158

Table 5 displays the Gibbs sampling and ICM estimates of the disturbance covariance matrix. Note the similarity between the two matrices.

5 Conclusion

A Bayesian factor analysis model was detailed in which available prior information either from substantive experts or previous experiments is quantified and incorporated into the inferences along with current data. An added feature of the Bayesian factor analysis model is that the there is no need to rotate the factor loading matrix. The rotation is automatically found. In addition, knowledge regarding the parameter values is allowed to accumulate as subsequent data is acquired. Available prior information regarding parameters was incorporated through generalized natural conjugate prior distributions which are rich enough to permit complete freedom of assessment.

References

[1] A. E. Gelfand and A. F. M Smith. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6:721–741, 1984.

[3] K. Hayashi. *The Press-Shigemasu Bayesian Factor Analysis Model With Estimated Hyperparameters.* PhD thesis, Department of Psychology, University of North Carolina, Chapel Hill, NC 27599, 1997.

[4] M. Kendall. *Multivariate Analysis.* Charles Griffin & Company LTD, London, second edition, 1980.

[5] D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34(1), 1972.

[6] A. O'Hagen. *Kendalls' Advanced Theory of Statistics, Volume 2B Bayesian Inference.* John Wiley and Sons Inc., New York, 1994.

[7] S. J. Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference.* Robert E. Krieger Publishing Company, Malabar, Florida, 1982.

[8] S. J. Press. *Bayesian Statistics: Principles, Models, and Applications.* John Wiley and Sons, New York, 1989.

[9] S. J. Press and K. Shigemasu. Bayesian inference in factor analysis. In *Contributions to Probability and Statistics*, chapter 15. Springer-Verlag, 1989.

[10] S. J. Press and K. Shigemasu. Bayesian inference in factor analysis-Revised. Technical Report No. 243, Department of Statistics, University of California, Riverside, CA 92521, May 1997.

[11] D. B. Rowe. *Correlated Bayesian Factor Analysis.* PhD thesis, Department of Statistics, University of California, Riverside, CA 92521, December 1998.

[12] D. B. Rowe. On estimating the mean in Bayesian factor analysis. Social Science Working Paper 1096, Division of Humanities and Social Sciences, Caltech, Pasadena, CA 91125, July 2000a.

[13] D. B. Rowe. Incorporating prior knowledge regarding the mean in Bayesian factor analysis. Social Science Working Paper 1097, Division of Humanities and Social Sciences, Caltech, Pasadena, CA 91125, July 2000b.

[14] D. B. Rowe and S. J. Press. Gibbs sampling and hill climbing in Bayesian factor analysis. Technical Report No. 255, Department of Statistics, University of California, Riverside, CA 92521, May 1998.

[15] D. M. Shera. *Bayesian Factor Analysis.* PhD thesis, Harvard University, School of Public Health, Boston, MA 02115, March 2000.